DOCUMENT RESUME

ED 418 603 FL 025 188

AUTHOR Davidson, Fred; Lynch, Brian K.
TITLE A Criterion-Referenced Viewpoint on Standards/Cutscores in Language Testing.
PUB DATE 1998-03-00
NOTE 54p.; Revised version of a paper presented at the Annual Meeting of the American Association for Applied Linguistics (Seattle, WA, March 14-17, 1998).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Academic Standards; *Criterion Referenced Tests; *English; Evaluation Criteria; Higher Education; Language Skills; *Language Tests; Plagiarism; *Second Languages; *Test Construction; Testing
IDENTIFIERS *International English Language Testing System; *University of Illinois Urbana Champaign

ABSTRACT

"Standard" is distinguished from "criterion" as it is used in criterion-referenced testing. The former is argued to refer to the real-world cutpoint at which a decision is made based on a test's result (e.g., exemption from a special training program). The latter is a skill or set of skills to which a test is referenced. However, criteria can relate to standards using a layer or layers of mediating descriptive information such as benchmark or level descriptors. Examples of this relationship are shown in two language tests, the International English Language Testing System and an experimental test of avoidance of plagiarism used at the University of Illinois at Urbana-Champaign. Clarity, consensus, and communication among members of the test development team are seen as critical to the clear articulation of the relationship between criterion-referenced test and its corresponding real-world standard or cutpoint. Contains 17 references. (Author/MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

A criterion-referenced viewpoint on standards/cutscores in language testing

Revised version of a Paper presented at the American Association for Applied Linguistics (AAAL)
Seattle, Washington, USA
March, 1998

ED 418 603

by

Fred Davidson, University of Illinois at Urbana-Champaign

Brian K. Lynch, University of Melbourne

Abstract

"Standard" is distinguished from "criterion" as it is used in criterion-referenced testing. The former is argued to refer to the real-world cutpoint at which a decision is made based on a test's result (e.g. exemption from a special training program). The latter is a skill or set of skills to which a test is referenced. However, criteria can relate to standards via a layer (or several layers) of mediating descriptive information such as benchmark or level descriptors. Examples of this relationship are given from two language tests: the International English Language Testing System and an experimental test of avoidance of plagiarism at the University of Illinois at Urbana-Champaign. Clarity, consensus and communication amongst a test development team are seen as critical to the clear relationship of a criterion-referenced test to its real-world standard or cutpoint.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Fred Davidson

BEST COPY AVAILABLE

FL025188

1. Introduction: on standards, criteria and other defintional matters

In this paper, we plan to accomplish several goals. First, we will detail some theoretical and definitional issues and cite some relevant language testing literature on the topic of standard-setting and criterion-referencing. Second, we will illustrate two real-world CR applications in which standard-setting figures keenly. The first example is historical and reviews some developments in the International English Language Testing Service (IELTS) exam in the late 1980s; for this part of our paper we held a real-time dialog with Caroline Clapham and Liz Hamp-Lyons at AAAL 1998, reported here. We are fortunate that they are here today for they were intensively involved with the IELTS at that time. Our second example will relate a new test under development at the University of Illinois at Urbana-Champaign. It is a multiple-choice examination of avoidance of plagiarism, and the particular relationship of that test to the existing UIUC ESL essay exam is quite relevant to the question of standard-setting. We will close with an appeal to several key concepts: that well-articulated CR tests can also articulate standards, that such articulation depends in turn upon a well-articulated test specification, that such specifications are best developed by group consensus among interested parties. That the real relationship between a CR test and standards (i.e. cutpoints) is usually mediated by several layers of descriptive and contextual information.

In order to give a criterion-referenced testing (CRT) perspective on standard setting, it is first necessary to clarify and distinguish between the terms "criterion" and "standard". The criterion, in CRT terms, is the behavior or skill that is being tested or assessed (Popham 1978; Hudson & Lynch 1984; Brown 1989). CRT demands a detailed formulation of this criterion, and our own work in this area has led us to conclude that

test specifications are a critical part of that formulation (Davidson & Lynch 1993; Lynch & Davidson 1994). The term criterion has other interpretations in the literature, most notably one that confuses it with a cut-score or standard.

To confuse matters further, the term "standard" has its own variety of interpretations. The International Language Testing Association's Task Force on Testing Standards (TFTS) reported three of the most frequently found definitions or uses of the term in its survey conducted in 1994-95 (TFTS 1995). Davidson, Turner, and Huhta (1998) offer the following elaboration of those three meanings:

1) a standard can refer to a guideline of good practice, for example, an important standard of educational tests is that their developers demonstrate evidence of test validity. This meaning equates 'standards' (in the plural) with a code of professional practice or set of professional guidelines which could cover all stages of test development, from intitial construction, through trialing, and on to operational use...
2) a standard can refer to an expected performance. First, it can refer to an expected level on a numerical scale at which some decision is made; for example, a score of 35 out of 50 on a written driver's licensing exam qualifies the applicant to take the behind-the-wheel portion of the test. Alternatively, it can refer to descriptions of behavior at one or many levels of performance, for example, 'At level two, examinees can perform simple spoken transactions in the foreign language, such as those typically involved in negotiation of daily shopping.'...
3) a standard can refer to a widely-accepted test of a given skill; for example, one could claim that TOEFL is a standard for assessment of English as a second/foreign language..." (Davidson, Turner and Huhta, 1998) p. 303

For the purposes of this discussion, we will focus on the second meaning for standard. In the case of "an expected level on a numerical scale at which some decision is made", the criterion is the ability or skill being measured, and the standard is the particular score that has been designated as the expected or required level of that ability. However, for "descriptions of behavior at one or many levels of performance" there is the potential to confuse standard with criterion. We would emphasize that it is the selection of a particular description of a level of performance as "the expected level... at which some decision is made" that makes it a standard. That is, the description of a criterion may be embodied in a scale which has a set of level descriptors. If, for example, one of those levels is designated as what is necessary for exempting from further language study then it becomes a standard. For instance, if a student needs a level 3 on a five-level descriptor scale, then the description of the criterion/criteria at that level become part of the standard. The point is that the "cutscore" notion of "standard" is actually external to the test itself and lies in the domain of contextual use of the test results.

A more specific example would be the recently developed ESL Standards for Pre-K--12 Students (TESOL 1997). What are refered to there as "goals", we would term criteria--descriptions of English language ability such as "using English to achieve academically in all content areas." These very general goal/criterion statements are formulated with a bit more detail through descriptors (e.g., "following oral and written directions, implicit and explicit"), and are further elaborated and supported by sample progress indicators and detailed classroom vignettes tailored to a specific grade level, ESL proficiency level, language of instruction, content area, and geographical location. What turns these into standards, for the purposes of our discussion, is the fact that

particular criteria--with their descriptors, progress indicators, and vignettes--are designated as what is expected for a particular "grade-level cluster" (e.g., K-3).

The standard setting exercise, then, is one of selection. Hudson (1986) discusses this process in relation to mastery testing, pointing out that such testing relies upon the validity of the standard or cut-score, and that this validity has generally been called into question because of its apparent, or perhaps essential, arbitrariness (on this point see also Shepard 1984). Ultimately, in most standard setting procedures, experts are asked for a judgment about where the standard should be set. Standard setting thus seems to be a subjective methodology, which leads to the concern about arbitrariness. There are methods, however, for making this exercise of judgment systematic and informed (Popham 1978; Hambleton 1980; Shepard 1984). Furthermore, Hudson argues for the ability of standard setting to work hand in hand with the development and refinement of CRTs and result in clearer articulations of what is being tested. He points out that the process of developing CRTs "often brings out differences and disagreements not previously considered to be differences among instructors, administrators and materials developers" (p. 264-5), and that this can lead to a clarification of curriculum goals and the instructional processes for reaching them. He also ends with the interesting observation that "the standard should appear valid to those who are not testing specialists." (p. 269) We concur with his points about consensus, clarity, and communication.

The use of expert judgment in standard setting begs the question of the existence of appropriate experts to make these decisions. In the field of language testing, Powers and Stansfield (1985) used native English speaking nurses and patients as expert judges (assessing the performances of non native English speaking nurses) in their standard

setting study for the Test of Spoken English. Lumley, Lynch and McNamara (1994) compared medical doctors and ESL specialists as expert judges in their investigation of the standard for the Occupational English Test (assessing immigrant, non native English speaking health professionals as part of their registration for practice in Australia). Lumley et al., similar to Hudson's claim, found that "in the process of conducting a standard setting investigation, information of relevance to test validity can sometimes surface" (p. 38)--in this case, they found that the expert judges (medical doctors) offered feedback on the authenticity and representativeness of the test tasks. They also found that, although testing technologies such as generalisability theory and item response theory can help the standard setting process, "there can be no purely technical solution to the problem of standard setting in this context." (p. 39) Their study provides a reliable range of scores, expressed in logit values, within which to locate the standard or cut-score; the selection of that cut-score becomes a political one which will favor one group (e.g., the immigrant health professionals) versus another (e.g., the medical establishment, and, perhaps, the patients).

Where does a CRT perspective on standard setting take us, then? When Clapham (this colloquium) suggests that it is easier to understand a standard in relation to "subjectively marked" language tests, we would argue that this is due to the criterion being specifically formulated, in terms of the descriptors used in the "subjective marking"--these give articulation to the particular level that is chosen as the standard. By contrast, the scores from an "objectively marked" test are not (usually) associated with detailed descriptions of what they are measuring. And even if there is a test specification that describes what is being tested, a particular score will not be associated with its own descriptor. We would argue for a specification of what the test taker is

expected to demonstrate that is developed in detail regardless of the label given to a particular test: NRM or CRM. This detailed description of the criterion results in the possibility of greater transparency in standard setting, in that it should be possible to link scale levels or score levels to meaningful expressions of the standard being selected. In the case of objectively marked tests, the detailed description of what the test items are designed to measure should itself assist expert judgment methods such as the Angoff procedure described by other presenters in this colloquium. Most standard-setting procedures (such as the Angoff approach) rely at some point on expert judgment. We are interested in formalizing that expertise in a manageable fashion.

In addition to the clarity and meaningfulness of the standard, there is the question raised above in relation to the Powers & Stansfield (1985) and Lumley et al. (1994) studies--who are the experts, who decides on the standards. This question is intimately linked to the notion of meaningfulness as well. We have argued that the "who decides" question is crucial to the test specification process, in particular, and the test development process, in general (Lynch & Davidson 1994), and our forthcoming work will examine this question more closely. This "who decides" issue applies to the standard setting process as well. We would further argue that with rare exception, the same team of people should decide on both the specification of the criterion and the designation of the standard. To enhance validity, this team of people needs to represent a variety of expert knowledges, not just testing expertise. Again, consensus, clarity and communication should be the orders of the day.

2. CR and Standard-Setting: an established example

We will now proceed to our first example of a CR standard setting situation: the IELTS of some eight or ten years ago.

Clapham (1996) reports on a study on the effect of background knowledge on performance in the International English Language Testing System (IELTS). This is an assessment battery used worldwide for admission to tertiary education. Her study concerns the version of the IELTS which was in force in the late 1980s and early 1990s. That test provides an interesting example of criterion-referenced standards in action. Furthermore, as Clapham was at this colloquium and to take advantage of her extensive experience with this test, we suggested to her that we discuss this test in our paper. We are grateful that she agreed. Our discussion was a live question-and-answer with her, concerning issues about CR standard-setting in that version of the IELTS.

Before we begin, we must emphasize one important point. The IELTS has changed significantly since that time period. Most particularly, its test specifications--a key component of CR views of standards--have been significantly revised. Our discussion and dialog with her today will be strictly historical.

We will focus on the IELTS writing test for Module C (Arts and Social Science) students. Clapham (1996) details both the general use of that version of the IELTS and its specific-field modular design. As we proceed, we must again emphasize that this is a non-operational test specification. Of particular interest is her discussion (ca. page 72) of how the IELTS test specifications were written. That involved a feedback-laden, cyclical evolution of consensus among IELTS team members and interested parties, and we

contend that such a process is essential to CR test development, including a CR approach to standard-setting.

Following are the questions posed to Clapham during our presentation. The questions are based on both the 1996 book and on the IELTS specifications and rating materials (band descriptors) which Caroline provided us for this paper (see Appendix 1). At Caroline's suggestion, we also included Liz Hamp-Lyons in the discussion. Hamp-Lyons, also present in our AAAL98 colloquium, was a consultant on the development of the IELTS writing module.

(1) Who wrote the test specifications?

(2) How were they developed (revised, redrafted)?

(3) Were the IELTS band level descriptors used to select or define the tasks, e.g. those shown under "academic tasks" such as those shown below? If so, how? Is the reverse also true -- did the IELTS specifications help to revise or change the band descriptors? Following are some sample academic tasks from the writing specification for Section 2 (from Appendix 1, below):

b) <u>Academic Tasks</u>

The test should sample the candidates' ability to perform the following tasks (not necessarily in isolation):

(i) Organising and presenting data

(ii) Listing the stages of a procedure

(iii) Describing an object or event or sequence of events

(iv) Explaining how something works

(v) Presenting the solution to a problem

(vi) Presenting and justifying an opinion, assessment or hypothesis either directly or by implication

(vii) Comparing and contrasting evidence, opinions, implications and hypotheses

(viii) Arguing a case

(ix) Evaluating and challenging ideas, evidence and argument

(4) In the development of the test specifications, was there a particular discussion of standards in this sense: was there discussion of the link between the band levels and the real world consequences of an individual being assigned to one band level or another? For example, did you say things like: "You know, we have to build X [some skill/task] into the spec because X is part of the band level 7 descriptor and students need to be at band 7 to be able to survive in a university"? For instance, we note with interest that the same test specification states a pre-requisite band level--that is, it assumes "The primary focus for writing in this test should be in the range of Bands 5, 6, and 7". Why? Did those three bands (or one of them) constitute a salient score usage decision cutpoint or standard?

Discussions with Clapham and Hamp-Lyons at the AAAL98 presentation revealed that the key feature guiding the IELTS writing specifications and assessment tasks was the history of the exam. The IELTS was an evolution of the earlier "ELTS" (English Language Testing Service) examination. Scoring bands, test specifications, and assumptions about the test tasks were designed to reflect back on that earlier test. Hence, there was no real discussion (during the writing of the IELTS specs) of real-world

standards. Rather, test developers, consultants, and interested parties were concerned with pegging the new exam to the old.

This raises an interesting question: to what extent can a test be said to represent an external standard if the test is (itself) not designed with that standard in mind? If, instead, the test is designed to match its earlier versions or predecessors, and if those earlier versions had some link to real-world decision standards, do subsequent versions of the test inherit that real-world link? To adequately address this issue would probably require a more lengthy historical narrative of the creation of the entire history of an exam--in all its generations and versions.

3. CR and Standard-Setting: an example under development

We would now like to describe another CR-type standard setting. In this next case, our example is an ongoing test development problem in which the nature of the "standard" is under active discussion. Unlike the IELTS example previously discussed, the next example we describe is a criterion-referenced standard-setting problem that is very much in the present tense--it is decidedly NOT yet resolved.

Recently, the ESL program at the University of Illinois at Urbana-Champaign (UIUC) has begun a trial of a new component to its ESL Placement Test (EPT). The EPT is used as a follow-up assessment to the Test of English as a Foreign Language (TOEFL) to help determine whether additional ESL courses are needed once a new international student arrives at UIUC. The current EPT is comprised of an oral interview, a multiple-choice exam of English usage, and an essay based on a video mini-lecture and reading text. Our

discussion today concerns the essay, for which the specification is given as Appendix 2, below.

The essay is the primary deciding evidence for placement into or exemption from the ESL writing courses at UIUC. It is a holistically-scored writing sample. Raters are taken from the ranks of experienced instructors in the course sequences. The holistic rating descriptors--known as benchmarks--have evolved and continue to be refined with each major EPT season. There are actually two sets of benchmarks, one for incoming undergraduate and one for incoming graduate students. Both sets of operational holistic benchmarks are given in Appendix 3.

The ESL writing course curriculum includes attention to source-based writing, citation and paraphrasing skills. Increasingly, ESL writing instructors at UIUC report that plagiarism is a problem that helps to distinguish whether a student is ready to exit the writing course stream and join his or her mainstream peers. Furthermore, discussion with various departments and faculty members at UIUC indicates that plagiarism is a worry shared across the campus as well. This concern with source-based writing, citation conventions, paraphrasing, and avoidance of plagiarism is reflected in various ways and at various points in the holistic essay benchmark/descriptors (Appendix 3). We should also note that the componential rating guidelines used within the service courses for progress testing also contain notable discussion of plagiarism (Appendix 4). Taken together, the intake benchmarks (grad and undergrad) and the composition scale indicate that avoidance of plagiarism is a marker of high-level ESL writing performance at UIUC. That is, in order to exempt from ESL writing courses (at UIUC) or in order to pass up and out of those courses, a student had better not plagiarize. Clearly, avoidance of plagiarism

is not the only skill needed for exemption and exit, but it seems to be a particularly salient one.

Those anti-plagiarism skills form the "mandate" for a new experimental component on the UIUC EPT, which is euphemistically called "the plagiarism test". This test is a collection of multiple-choice items that could be phased into the present multiple choice test of English usage, which UIUC test developers wish to improve because it has a number of items displaying poor statistical quality and in need of replacement. This past January, they ran an experimental EPT "caboose" of plagiarism items with the Spring 1998 EPT intake group--that is, they added a set of zero-stakes plagiarism questions at the end of the EPT test session. This plagiarism test is derived from a test specification, which includes sample items and a revised rating rubric grid derived from the componential progress grid noted above and which was used in the analysis of the caboose (all of these materials are given as Appendix 5 below). This "plagiarism spec" has undergone extensive revision in a language testing course, in independent graduate student projects, and as the result of small-scale piloting prior to the recent trial caboose (see Lynch and Davidson, 1994, Figure 1). It is important to note that the specification requires--and the caboose included--a mini-lesson on avoidance of plagiarism as part of the exam.

There has been active debate in the L2 writing literature about the relevance of plagiarism to ESL writing. This debate has raised both methodological and substantive issues of deep concern; in this regard see Deckert 1993 and 1994 and Pennycook 1994 all published in The Journal of Second Language Writing (the UIUC Plagiarism Test was influenced by Deckert's 1993 work). In part due to that literature, language testing doctoral students, graduate student course instructors, faculty--that is, the language

testing "plagiarism" team at UIUC are frankly skeptical that this test is going to work. A chief suspicion is that source-based writing is too variable even within a writing community to derive common guidelines about how to avoid it in writing. Another suspicion is that the particular skills in the experimental items are overly detailed and culturally bound. There is also a worry that the trial test is just too difficult.

Nonetheless, the UIUC team perceives plagiarism to be part of its testing mandate. This perception comes from communication with the ESL writing instructors, who themselves must wear two hats: they teach the courses and also rate the intake essay on the EPT. These teachers consistently report that plagiarism--more accurately, its avoidance--is a key component to exemption from ESL writing instruction. According to the teachers, plagiarism seems to "mark" many essays for students who do NOT exempt from ESL writing instruction, and for those who do place into the ESL writing course sequence, avoidance of plagiarism is one of several key components of successful completion of the instruction. Again, we can note this concern is evident in the holistic EPT essay rating scales/benchmarks and in the componential progress testing grid.

On the one side is the mandated concern about plagiarism in the ESL service courses at UIUC. On the other is the debate in the literature and the consequent skepticism of many of those there who are presently crafting this test. There is need to resolve this tension. The UIUC team needs to determine if avoidance of plagiarism is in fact part of the skills necessary to be exempt from ESL writing classes at our campus. We do not wish to claim that plagiarism is (or for that matter is not) a valid standard of ESL college writing assessment (though we note with interest that the AAAL98 presentation yielded an active audience discussion of the topic of plagiarism). Rather, we are interested in examining the process by which it being considered at the UIUC context. The potential

role of plagiarism in the UIUC ESL exemption standard is evolving through a discursive process among a team of interested parties. This role is shown schematically in Figure 1. The UIUC team are confident that they will arrive at a feasible decision about how to assess plagiarism in their test. It is distinctly possible that they will decide not to run a multiple-choice plagiarism test as part of the EPT at all. Or it is possible that the experimental plagiarism caboose will trigger significant changes in the holistic benchmark scales and/or in the componential progress grid. Or it is possible that UIUC might decide to include an m/c plagiarism test but alter the items (and the spec) in significant ways.

Regardless, UIUC testers and teachers will hopefully clarify the role that plagiarism presently has in the current holistic benchmarks, and to some extent, that is what they have been after all along. That role--how plagiarism plays into the rating of the EPT essay--is the avenue by which we reach the exemption and exit standards for ESL writing at UIUC.

4. Concluding remarks: the nature of CR standard setting

In this paper we have sketched a criterion-referenced perspective on standard-setting. We contend that well-articulated criterion-referenced (CR) tests can also articulate standards. Such CR articulation depends upon a sound test specification, and within the spec and its associated materials should be a descriptive textual linkage to show how the test relates to a real-world standard. We gave a specific model of this linkage as Figure 1 (for the UIUC context) above. A more generalized model is shown in Figure 2: the test and its specification are related to an external real-world decision cutpoint via the

mediation of some set of level descriptors/benchmarks. For example, in the IELTS case, our dialogue with Caroline and Liz indicated that the late 1980s / early 1990s IELTS test developers did not revise their perceptions of the external standards for IELTS score users. Rather, the new test (the IELTS) was patterned on the old (the ELTS), and further historical research would be needed to clarify the IELTS-external standard relationship implied by Figure 2. However, at UIUC, this relationship is more clear--if more controversial. As shown in Figure 1 (which is a special case of Figure 2), plagiarism is being investigated as a particularly salient skill to denote the real-world cutpoint/standard of exemption from ESL writing courses.

In both cases--and in well crafted CR testing in general--we would contend that test specifications are best developed by group consensus among interested parties. This consensus should include attention to real-world decision cutpoints/standards. The real relationship between a CR test and standards is usually mediated by several layers of descriptive and contextual information, which we have represented (in Figures 1 and 2) as descriptors/benchmarks. We adopt the attitude that "there can be no purely technical solution to the problem of standard setting" (Lumley, Lynch and McNamara, 1994: 39) in most CR-based standard-setting contexts. However, the notion of expert judgment (which is a key component of most standard setting protocols) needs to be given more attention and made more systematic. Instead of advocating the statistical formalization of that expert judgment, we seek a non-statistical description and formalization of consensus-building among members of a testcrafting team.

References

Brown, J. D. 1989. Improving ESL placement tests using two perspectives. TESOL Quarterly, 7, 239-260.

Clapham, Caroline. 1996. The Development of IELTS: A Study of the Effect of Background Knowledge on Reading Comprehension. Cambridge, UK: University of Cambridge Local Examinations Syndicate and Cambridge University Press. Volume 4 in the series: Studies in Language Testing.

Davidson, F. & Lynch, B. K. 1993. Criterion-referenced test development: a prolegomenon. In A. Huhta, K. Sajavaara, & S. Takala (Eds.) Language Testing: New Openings, Jyväskylä, Finland: University of Jyväskylä, Institute for Educational Research.

Davidson, F., Turner, C. E., & Huhta, A. 1998. Language testing standards. In C. Clapham (Ed.) The Encyclopedia of Language and Education Volume 7: Language testing and assessment. Dordrecht, Netherlands: Kluwer Academic Publishers.

Deckert, G. 1993. Perspectives on plagiarism from ESL tudents in Hong Kong. Journal of Second Language Writing 2(2), pp. 131-148.

---------. 1994. Author's response to Pennycook's objections. Journal of Second Language Writing 3(3), pp. 284-289.

Hambleton, R.K. 1980. Test score validity and standard setting methods. In R.A. Berk (Ed.), Criterion-referenced Measurement: State of the Art. Baltimore, MD: Johns Hopkins University Press, pp. 80-123.

Hudson, T. D. & Lynch, B. K. 1984. A criterion-referenced measurement approach to ESL achievement testing. Language Testing 1, 171-201.

Hudson, T. D. 1986. Mastery decisions in program evaluation. In R. K. Johnson (Ed.) The Language Curriculum, pp. 259-269. Cambridge: Cambridge University Press.

Lumley, T., Lynch, B. K. & McNamara, T. F. 1994. A new approach to standard-setting in language assessment. Melbourne Papers in Language Testing, 3(2), 19-39.

Lynch, B. K. & Davidson, F. 1994. Criterion-referenced language test development: linking curricula, teachers, and tests. TESOL Quarterly, 28, 727-743.

Pennycook, A. 1994. The complex contexts of plagiarism: a reply to Deckert. Journal of Second Language Writing 3(3), pp. 277-284.

Popham, W. J. 1978. Criterion-refenced Measurement. Englewood Cliffs, NJ: Prentice Hall.

Powers, D. E. & Stansfield, C. W. 1985. Testing the oral English proficiency of foreign nursing graduates. The ESP Journal, 4(1), 21-36.

Shepard, L. A. 1984. Setting performance standards. . In R.A. Berk (Ed.), Criterion-referenced Measurement: State of the Art. Baltimore, MD: Johns Hopkins University Press, pp. 80-123.

Task Force on Testing Standards (TFTS). 1995. Report of the Task Force on Testing Standards (TFTS) to the International Language Testing Association. Lancaster, UK: ILTA. Available online: http://www.surrey.ac.uk/ELI/ilta/ilta.html.

TESOL. 1997. ESL Standards for pre-K--12 students. Alexandria, VA: TESOL Publications.

Appendix 1: Excerpts from the "International English Language Testing System [IELTS]: Specifications for Module C (December, 1989)", "Academic Modules Profile Band Descriptors", and "Global Band Descriptors", provided by Caroline Clapham. **Not operational**

3. Section 2: Writing (45 minutes)

Test Focus

a) Band Levels
The primary focus for writing in this test should be in the ranges of Bands 5, 6, and 7. (See Writing Global Band Descriptors in Appendices 2 and 3)

b) Academic Tasks
The test should sample the candidates' ability to perform the following tasks (not necessarily in isolation):

(i) Organising and presenting data
(ii) Listing the stages of a procedure
(iii) Describing an object or event or sequence of events
(iv) Explaining how something works
(v) Presenting the solution to a problem
(vi) Presenting and justifying an opinion, assessment or hypothesis either directly or by implication
(vii) Comparing and contrasting evidence, opinions, implicatoins and hypotheses
(viii) Arguing a case
(ix) Evaluating and challenging ideas, evidence and argument

c) Audience
Appropriate audiences are:

(i) Professorial--e.g. supervisors, teachers, examiners
(ii) Professional--e.g. practitioners in the field, fellow students, clients
(iii) Personal--e.g. writing for own use

Stimulus Materials

a) Level
Where completion of the writing task depends on reading, the reading should not require proficiency greater than Band 5.

b) Texts
Stimulus material may be textual, diagrammatic, graphic, or photographic. Graphs and tables should be simple to interpret and be fully labelled. Texts must be realistic and in modern English, but may be authentic, modified or constructed.

c) Length
The time required to understand stimulus material should not be more that five minutes

Test Tasks

There should be two writing tasks, each of which should generate enough writing to provide sufficient information for the answer to be assigned to a Band Level. At least one of the tasks should draw on one or more of the reading passages in Section 1. One task may be based on stimulus material presented solely for the writing task.

For each of the tasks a set of guidelines is provided (see below) together with a template containing the basic rubric to appear in each version of the task.

a) Writing Task 1 (15 minutes)

This task involves information transfer or reprocessing. (See Template on Page 9.)

Length of answer: at least 100 words

Content: 1. content will be provided, either via the reading texts or via a specially provided text

2. The stimulus material should be related to the Arts and Social Sciences

Texts: 1. input text material, previously read or new, must be such that a student whose reading level is Band 5 can process it in less than 5 minutes.

2. The texts should not contain language structures that can be transferred into an answer, although key lexis may be transferred.

Mode: 1. description/narration focussing on process, i.e., the following academic tasks:

(i) Listing the stages of a procedure
(ii) Describing an object/event/squence
(iii) Explaining how something works
(iv) Organising and presenting data

2. (short) essay, report

Audience: varied

Marking criteria: 1. the assessor should refer to the Assessment Guide

2. Item writers should refer to the Assessment Guide but be primarily guided by the criteria in the Template.

Level of Difficulty: Bands 5 to 7

b) Template for Writing Task 1

(See Specimen Materials Booklet for an example of Writing Task 1.)

You should spend no more than 15 minutes on this task.

[Short summary of topic covered in the source material, or reference to what the diagram, graph or table shows.]

Task: [The task should be as simply worded as possible.]

You may use your own knowledge and experience in addition to the [diagram/graph/table]

Make sure your description is:

1. Relevant to the question, and

2. Well organised.

You should write at least 100 words

[Heading for diagram, table, etc. where necessary]

[diagram, graph, table etc. where necessary]

Note: material in square brackets is to be provided by item writer.

c) Writing Task 2 (30 minutes)

This task involves analysis and synthesis, and calls on personal experience, knowledge and views. (See Template on Page 11.)

Length of answer: at least 150 words

Content: 1. topic extracted from material in the source texts; should be something referred to, not necessarily in detail.

2. Topic should generally appear in more than one text

3. Level of specificity is only slight: essentially, topics should not depend on any degree of background knowledge in a disciplinary area; the aim is face validity not content or construct validity.

Texts: 1. candidates are asked to support their argument by citing relevant evidence from the Reading Passages. They less they already know at this point the more likely they are to turn back to the texts.

2. The structure of the actual question SHOULD make plagiarism (a) unlikely (b) so unsuitable that a plagiarised answer would be penalised on task fulfillment grounds.

Mode: 1. essay

2. argument (which includes a personal element) with argument defined as necessitating some consideration of opposing views, i.e., the following academic tasks:

(i) explaining why something is the case
(ii) presenting and justifying an assessment, hypothesis or opinion either directly or by implication
(iii) comparing and contrasting evidence, opinions, hypotheses and implications
(iv) presenting the solution to a problem
(v) arguing a case
(vi) evaluating and challenging ideas, evidence and arguments

Audience: a university teacher

Marking Criteria: 1. the assessor should refer to the Assessment Guide

2. item writers should refer to the Assessment Guide but be primarily guided by the criteria in the Template

Level of Difficulty: Bands 5 to 7

d) Template for Writing Task 2

(See Specimen Materials Booklet for an example of Writing Task 2.)

You should spend no more than 30 minutes on this task.

Task:

Write an essay for a university teacher on the following topic:

[Title of Essay]

In writing your essay, make sure that:

1. the essay is well organised
2. your point of view is clearly expressed, and
3. your argument is supported by relevant evidence from the Reading Passages

NOTE: do not copy word for word from the Reading Passages.

You should write at least 150 words.

SPACE FOR NOTES

Note: Material in square brackets is to be provided by the item writer.

ACADEMIC MODULES–GLOBAL BAND DESCRIPTORS

QUESTION 1

9. This is an answer which fulfills the task in a way which the reader finds completely satisfactory. The message can be followed effortlessly. Coherence and cohension are so skillfully managed that they attract no attention. A wide range of sentence structures is used accurately and appropriately.

8. This answer does not fully achieve the level of a 9 in either task fulfillment or coherence or cohesion, but the range of sentence structures used is good, and is well controlled for accuracy and appropriacy.

7. This is a satisfactory answer which generally addresses the task more relevantly, appropriately and accurately although the reader notices that it could be more fully developed. The message can be followed throughout, and usually with ease. Information is generally arranged coherently, and cohesion within and between sentences is satisfactorily managed. A satisfactory range of sentence structures occurs, and there are only occasional minor flaws in the control of sentence structure.

6. This is a mainly satisfactory answer which generally addresses the task. The reader notices some irrelevant, inappropriate or inaccurate information but only in areas of minor importance. There may be minor details missing. The message can be followed throughout. Information is generally arranged coherently, but cohesion within and/or between sentences may be faulty, with misuse, overuse or omission of cohesive devices. Sentence structures are generally inadequate, but the reader may feel that control is achieved by the use of a restricted range of structures, or, in contrast, that the use of a wide variety of structures is not marked by the same level of skill and accuracy.

5. This is an adequate answer but the inclusion of irrelevant, inappropriate or inaccurate material in key areas detracts from its fulfillment of the task. There may be some details missing. The message can generally be followed although sometimes only with difficulty. Both coherence and cohesion may be faulty. There is a limited range of sentence structures, and the greatest accuracy is achieved on short, simple sentences. Errors in such areas as agreement of tenses or subjects and verbs are noticeable.

4. This answer attempts to fulfill the task but is prevented from doing so adequately by considerable amounts of irrelevance, inappropriacy or inaccuracy. There may be some details missing. The message is difficult to follow. Information is not arranged coherently and cohesive devices are inadequate or missing. Limited control of sentence structures, even short, simple ones is evident. Errors in such areas as agreement of tenses or of subjects and verbs cause severe strain for the reader.

3. The seriousness of the flaws in this answer make it difficult to judge in relation to the task. The message cannot be followed. Neither coherence nor cohesion are apparent. Control of sentence structure is evident only occasionally and errors predominate.

2. This answer does not reach the level of 3 in task fulfillment. There is no recognizable message. There is little or no evidence of control of sentence structure.

1. The writing appears to be by a virtual non-writer, containing no assessable strings of English writing. If an answer is wholly or almost wholly copied from the source materials it is scored in this category.

0. Should only be used where a candidate did not attend or did not attempt this question in any way.

BEST COPY AVAILABLE

ACADEMIC MODULES GLOBAL BAND DESCRIPTORS

QUESTION 2

9. The reader finds the essay completely satisfactory. A point of view is presented and developed, either arguing for and supporting one position or considering alternative positions by presenting and discussing relevant ideas and evidence. The argument proceeds logically through the text with a clear progression of ideas. There is plentiful material. A wide range of vocabulary is used appropriately. The reader sees no errors in word formation or spelling. A wide range of sentence structures is used accurately and appropriately.

8. This answer does not fully achieve the 9 level in communicative quality, arguments, ideas and evidence. There is a good range of appropriate vocabulary. The reader sees no signficant errors in word formation or spelling. The range of sentence structures used is good, and is well controlled for accuracy and appropriacy.

7. The reader finds this a satisfactory essay which generally communicates fluently and only rarely causes strain. A point of view is presented, although it may be unclear at times whether a single position is being taken or alternative positions being considered. The argument has a clear progression overall although there may be minor isolated problems. Ideas and evidence are relevant and sufficient but more specific detail may seem desirable. The range of vocabulary is fairly good and vocabulary is usually used appropriately. Errors in word formation are rare and, while spelling errors do occur, they are not intrusive. A satisfactory range of sentence structures occurs and there are only occasional, minor flaws in the control of sentence structure.

6. The reader finds this is a mainly satisfactory essay which communicates with some degree of fluency. Although there is sometimes strain for the reader, control of organisational patterns and devices is evident. A point of view is presented although it may be unclear whether a single position is being taken or alternative positions are being considered. The progression of the argument is not always clear, and it may be difficult to distinguish main ideas from supporting material. The relevance of some ideas or evidence may be dubious and some specific support may seem desirable. The range of vocabulary sometimes appears limited as does the inappropriacy of its use. Minor limitations of, or errors in, word choice sometimes intrude on the reader. Word formation and spelling errors occur but are only slightly intrusive. Sentence structures are generally inadequate but the reader may feel that control is achieved by the use of a restricted range of structures or, in contrast, that the use of a wide variety of structures is not marked by the same level of structual accuracy.

5. This is an essay which often causes strain for the reader. While the reader is aware of an overall lack of fluency, there is a sense of an answer which has an underlying coherence. The essay introduces ideas although there may not be many of them or they may be insufficiently developed. Arguments are presented but may lack clarity, relevance, consistency or support. The range of vocabulary and appropriacy of its use are limited. Lexical confusion and incorrect word choice are noticeable. Word formation and spelling errors may be quite intrusive. There is a limited range of sentence structures and the greatest accuracy is achieved in short, simple sentences. Errors in such areas as agreement of tenses or subjects and verbs is noticeable.

4. The essay attempts communication but meaning comes through only after considerable effort by the reader. There are signs of a point of view but main ideas are difficult to distinguish from supporting material and the amount of support is inadequate. Such evidence and ideas as are presented may not be relevant. There is no clear progression to the argument. The range of vocabulary is often inadequate and/or inappropriate. Word choice causes serious problems for the reader. Word formation and spelling errors cause severe strain for the reader. Limited control of sentence structures, even short, simple ones, is evident. Errors in such areas as agreement of tenses, or of subjects and verbs cause severe strain for the reader.

3. The seriousness of the problems in this essay prevent meaning from coming through more than spasmodically. The essay has few ideas and no apparent development. Such evidence and ideas as are presented are irrelevant. There is little comprehensible point of view or argument. The reader is aware of gross inadequacies of vocabulary, word forms and spelling. Control of sentence structre is evident only occasionally and errors predominate.

2. The writing displays no ability to communicate. There is evidence of one or two ideas without development. The reader sees no control of word choice, word forms and spelling. There is little or no evidence of control of sentence structure.

1. The writing appears to be by a virtual non-writer, containing no assessable strings of English writing. If an answer is wholly or almost wholly copied from the source materials it is scored in this category.

0. Should only be used where a candidate did not attend or did not attempt this question in any way.

ACADEMIC MODULES PROFILE BAND DESCRIPTORS
QUESTION 1

	TASK FULFILLMENT	COHERENCE & COHESION	SENTENCE STRUCTURE
9	This is an answer which fulfills the task in a way which the reader finds completely satisfactory.	The message can be followed effortlessly. Coherence and cohesion are so skillfully managed that they attract no attention.	A wide range of sentence structures is used accurately and appropriately.
8	Does not fully achieve the level of a 9.	Does not achieve the level of a 9.	The range of sentence structures used is good, and is well controlled for accuracy and appropriacy.
7	This is a satisfactory answer which generally addresses the task relevantly, appropriately, and accurately, although the reader notices that it could be more fully developed.	The message can be followed throughout and usually with ease. Information is generally arranged coherently and cohesion within and between sentences is satisfactorily managed.	A satisfactory range of sentence structures occurs and there are only occasional minor flaws in the control of sentence structure.
6	This is a mainly satisfactory answer which generally addresses the task. The reader notices some irrelevant, inappropriate or inaccurate information, but only in areas of minor importance. There may be minor details missing.	The message can be followed throughout. Information is generally arranged coherently but cohesion within and/or between sentences may be faulty with misuse, overuse or omission of cohesive devices.	Sentence structures are generally inadequate, but the reader may feel that control is achieved by the use of a restricted range of structures or, in contrast, that the use of a wide variety of structures is not marked by the same level of accuracy.
5	This is an adequate answer but the inclusion of irrelevant, inappropriate or inaccurate material in key areas detracts from the fulfillment of the task. There may be some minor details missing.	The message can generally be followed although sometimes only with difficulty. Both coherence and cohesion may be faulty.	There is a limited range of sentence structures and the greatest accuracy is achieved on short, simple sentences. Errors in such areas as agreement of tenses or subjects and verbs are noticeable.
4	This answer attempts to fulfill the task but is prevented from doing so adequately by considerable amounts of irrelevance, in-appropriacy or inaccuracy. There may be some details missing.	The message is difficult to follow. Information is not arranged coherently, and cohesive devices are inadequate or missing.	Limited control of sentence structures, even short, simple ones, is evident. Errors in such areas as agreement of tenses or of subjects and verbs can cause severe strain for the reader.
3	The seriousness of the flaws in this answer make it difficult to judge in relation to the task.	The message cannot be followed. Neither coherence nor cohesion are apparent.	Control of sentence structure is evident only occasionally and errors predominate.
2	Does not reach the level of a 3.	There is no recognizable message.	There is little or no evidence of control of sentence structure.
1	The writing appears to be by a	Not applicable.	Not applicable.

virtual non-writer, containing no assessable strings of English writing. If an aswer is wholly or almost wholly copied from the source materials, it is scored in this category.

0. Should only be used where a candidate did not attend or did not attempt this question in any way.

ACADEMIC MODULES PROFILE BAND DESCRIPTORS
QUESTION 2

	COMMUNICATIVE QUALITY	ARGUMENTS, IDEAS AND EVIDENCE	WORD CHOICE, FORM, AND SPELLING	SENTENCE STRUCTURE
9.	The reader finds the essay completely satsifactory	A point of view is presented and developed, either arguing for and supporting one position or considering alternative positions by presenting and discussing relevant evidence. The argument proceeds logically through the text, with a clear progression of ideas. There is plentiful material.	A wide range of vocabulary is used appropriately. The reader sees no errors in word formation or spelling.	A wide range of sentence structures is used accurately and appropriately.
8.	Does not fully achieve the level of a 9.	Does not achieve the level of a 9.	There is a good range of appropriate vocabulary. The reader sees no significant errors in word formation or spelling.	The range of sentence structures is good, and is well controlled for accuracy and appropriacy.
7.	The reader finds this a satisfactory essay which generally communicates fluently and only rarely causes strain.	A point of view is presented although it may be unclear at times whether a single position is being taken or alternative positions being considered. The argument has a clear progression overall although there may be minor isolated problems. Ideas and evidence are relevant and sufficient but more specific detail may seem desirable.	The range of vocabulary is fairly good and vocabulary is usally used appropriately. Errors in word formation are rare, and, while spelling errors do occur, they are not intrusive.	A satisfactory range of sentence structures occurs and there are only occasional minor flaws in the control of sentence structure.
6.	The reader finds this a mainly satisfactory essay which communicates with some degree of fluency. Although there is sometimes strain for the reader, control of organisational patterns and devices is evident.	A point of view is presented although it may be unclear whether a single position is being taken or alternative positions are being considered. The progression of the argument is not always clear and it may be difficult to distinguish main ideas from sup-	The range of vocabulary sometimes appears limited as does the appropriacy of its use. Minor limitations of, or errors in, word choice sometimes intrude on the reader. Word formation and spelling errors occur but are only slightly intrusive.	Sentence structures are generally adequate, but the reader may feel that control is achieved by the use of a restricted range of structures or, in contrast, that the use of a wide variety of structures is not marked by the same level of structural

BEST COPY AVAILABLE

		porting material. The relevance of some ideas or evidence may be dubious and more specific support may seem desirable.		accuracy.
5.	This is an essay which often causes strain for the reader. While the reader is aware of an overall lack of fluency, there is a sense of an answer which has an underlying coherence.	The essay introduces ideas athough there may not be many of them or they may be insufficiently developed. Arguments are presented but may lack clarity, relevance, consistency or support.	The range of vocabulary and the appropriacy of its use are limited. Lexical confusion and incorrect word choice are noticeable. Word formation and spelling errors may be quite intrusive.	There is limited range of sentence structures and the greatest accuracy is achieved on short, simple sentences. Errors in such areas as agreement of tenses or subjects and verbs are noticeable.
4.	The essay attempts communication but meaning comes through only after considerable effort by the reader.	There are signs of a point of view but main ideas are difficult to distinguish from supporting material and the amount of support is inadequate. Such evidence and ideas as are presented may not be relevant. There is no clear progression to the argument.	The range of vocabulary is often inadequate and/or inappropriate. Word choice causes serious problems for the reader. Word formation and spelling errors cause severe strain for the reader.	Limited control of sentence structures, even short simple ones, is evident. Errors in such areas as agreement of tenses or of subjects and verbs cause severe strain for the reader.
3.	The seriousness of the problems in this essay prevent meaning from coming through more than spasmodically.	The essay has few ideas and no apparent development. Such evidence and ideas as are presented are ir-relevant. There is little comprehensible point of view or argument.	The reader is aware of gross inadequacies of vocabulary, word forms and spelling.	Control of sentence structure is evident only occasionally and errors predominate.
2.	The writing displays no ability to communicate.	There is evidence of one or two ideas without development.	The reader sees no control of word choice, word forms and spelling.	There is little or no evidence of control of sentence structure.
1.	The writing appears to be by a virtual non-writer, containing no assessable strings of English writing. If an answer is wholly or almost wholly copied from the source materials it is scored in this category.	Not applicable.	Not applicable.	Not applicable.

0. Should only be used where a candidate did not attend or did not attempt this part of this question in any way.

BEST COPY AVAILABLE

Appendix 2: UIUC ESL Placement Test Video-Reading Essay Exam: Test Specification as of February 1998.

The EPT Specification for Video-Reading Based Academic Essays

Origin / Comments:
[Nb. Slight additional changes made from Angie Liu's (1997) dissertation in Feb 1998 to accurately cross-reference the holistic EPT benchmarks. -FD]
This version of the *EPT Specification for Video-Reading Based Academic* Essays has evolved from the past five versions. The five earlier versions were constructed on the basis of a criterion-referenced-based specification format developed by Popham (1978), as well as the principals under the criterion-referenced language test development (CRLTD), presented at the 1994 Language Testing Colloquium in Washington, D.C. (Davidson, Lynch, Cho, & Larson, 1994; Lynch & Davidson, 1994). Other contributors of the earlier versions of the EPT test specification include Prof. Susan Larson from the department of civil engineering and the former-EPT reserarch assistant, Dongwan Cho. The current version of the *EPT Specification for Video-Reading Based Academic Essays* represents both evolution and change in the original conceptual framework of the performance-based, authentic, academic writing tasks. And, the change in this version was influenced by prompt evaluators (David Broersma, Anna Kasten, Gene Hennigh, and Volker Hegelheimer), the instructional technology specialist (Tim Genvey), lecturing professors (Prof. Robert Wengert, Prof. Larry Debrock, and Prof. Molly Mack), the EPT supervisor (Prof. Fred Davidson), and the EPT research assistant (Angie Liu).

Related Specifications, if any: EPT Multiple-Choice Test of Awareness of Plagiarism

GD: General Abilities / Skills Being Tested
In order to demonstrate the ability in comprehending and producing academic English essays accepted in most U.S. universities, examinees need to successfully complete the task of integrating information from different sources (i.e. the academic / non-technical lecture in a videotape and the reading text of the same theme) and presenting it in a general writing format (i.e. introduction, body and conclusion).
The specific abilities / skills being tested are:

a) obtaining information on a given theme from different source channels, for instance, listening to lectures and reading pertinent texts.
b) understanding main ideas and being able to distinguish them from minor ones.
c) taking notes while listening to academic lectures and using the notes to develop the subsequent writing task.
d) integrating and synthesizing the information given and presenting it in a general writing format -- namely, introduction, body and conclusion.
e) writing in one's own words, paraphrasing the information given.
f) developing a main idea about the topic and support that idea with information from the academic lecture and the reading text.

SI: Directions for the Writing Task
Generally speaking, examinees will receive instructions as to what types of prompts they would expect, what they should do to accomplish the writing task successfully and how their essays will be evaluated.

Specific Procedures:

1. Show the video with the academic lecture (7-11 minutes) to examinees.
2. Ask examinees to read the reading text provided and tell them to start the writing task whenever they are ready. (*Time for reading the text depends on individual examinees).

Directions

BEST COPY AVAILABLE

In a moment, you will watch a videotape entitled "_____?". The videotape is about ___ minutes long and will be played only once. While watching the videotape, you can take notes on the back of this instruction sheet. Please note that your notes will not be graded. After watching the videotape, read the article which will be provided to you. When you are ready, start writing a 1 to 2 page essay in this booklet based on the information in the videotape and the reading article. In your writing, you should develop a main idea about the topic and support that idea with information from the videotape and the article. The following criteria will be used to grade your essay:

a) Your essay should have a clear introduction, body and conclusion.
b) The ideas within your essay should be explicitly connected.
c) Your ideas should be supported with the evidence from both the videotape and the article.
d) Your essay should be written in your own words. Don not reproduce directly the videotape and the article in your essay.
e) Your essay should demonstrate the use of standard grammatical conventions.

You will have 50 minutes to read the article and to write the essay after the video is stopped. The time left will be put down on the blackboard every ten minutes.

PA: Characteristics of the Stimuli

Video Lecture:

a) The level of information should be general and academic, but not too technical.
b) The content should be culturally appropriate.
c) A title should be provided for the video-lecture and appear at the beginning of the video.
d) The length of the video-lecture is permitted to range from 7 to 11 minutes.
e) University professors without particular accents are qualified candidates to serve as speakers of the video lecture. Based on individual professor's specialty and constrains of the VRESSAY (e.g., the length of the lecture), the invited speaker will develop the necessary script of the lecture and deliver it at a natural speed.
f) Audience are allowed to be present at the lecture; however, no lecturer-audience interaction is permitted.
g) To ensure the quality of the videotape, the shooting should take place in a high-quality studio available from campus resources.
h) Preferably (not required), there is a summary section provided at the end of the video lecture.
i) Preferably (not required), teaching aids such as pictures, graphs, realia, or notes on the blackboard are used to facilitate presenting the information in the video lecture. If teaching aids are used, the quality of them should be ensured so that examinees can actually take advantage of those aids to process the information.

Reading Text:

a) The level of the information should be general and academic, but not too technical.
b) The content should be culturally appropriate.
c) The length of the reading text is permitted to range from 600 to 1000 words.
d) It should discuss the same thematic topic as the video lecture.
e) It should contain information which is related to but different from those of the video lecture (e.g. general vs. specific information; opposing viewpoints; theory vs. application, simplified view vs. complicated view, less information vs. more information ... etc.)
f) It can be selected from authentic college textbooks, journal articles of non-technical nature, prestigious magazines or newspapers. In this case, the reference citation should not appear in the reading text for the sake of test security, but should be noted in test archives.
g) It can be written, rewritten, or edited by native speakers of English on the basis of authentic materials to strengthen its link with the video lecture or to adjust the level of readability.

RA: Descriptions of the Expected Writing Response and Format

Examinees will be given a test booklet which includes space for some background information, instructions for the writing task, a blank page for note-taking, as well as response pages for essay samples.

The grading system involves expert judgment by raters who have received training in using the UIUC EPT holistic essay scoring scale (known as "the benchmarks") and have experience in teaching ESL service courses in the academic level they are grading (e.g. teaching assistants for undergraduate ESL classes will evaluate only undergraduate examinees' essay samples. The same principle applies to graduate examinees.). At the beginning of each grading session, the raters will watch the video again and read the reading passage. The raters will then recalibrate their level-scales by discussing the match between the quality of essays and the level of ESL courses needed. Each essay will be graded blindly by two raters. The higher score will be used to place examinees into appropriate ESL courses if the rating discrepancy is within one level-score. A third rater will be asked to further evaluate the examinees' essays if the rating discrepancy is more than one level-score. The final placement will be made on the basis of the match between the third rater and either of the first two raters.

Specifically, the following criteria will be used to grade examinees' essays:

a) The essay should have a clear introduction, body and conclusion.
b) The ideas within the essay should be explicitly connected.
c) The ideas should be supported with evidence from both the videotape and the reading text.
d) The essay should be written in examinees' own words. Information can not be reproduced directly from the videotape or the reading text.
e) The essay should demonstrate the use of standard grammatical conventions.

SS: Supplementary Information

Following guidelines regarding the format of the materials used in the video-lectures are recommended by a video filming / editing specialist (i.e., Mr. Tim Genvey in the Office of Instructional Resources) to enhance the quality of the video productions:

a) There should be less than thirty characters in a line (including space) if text information is presented.
b) Maximally, seven lines are allowed on one page if text information is presented.
c) Do not leave out too much blank space on a page.
d) The information, including both texts and graphics, is better presented in landscape orientation.
e) Transparencies and slides are discouraged from being used because of the unwanted impact they will create on the screen (i.e. oftentimes they are too shining).
f) Hard-copy of the materials which are used to aid the presentation of the lecture, including text and graphics, are suggested to be used.
g) Hard-copy materials are recommended to be printed on blue powder paper.
h) Hard-copy pictures can not be shining.
i) If computer files are to be used, the text information is suggested to be formatted using the powerpoint software.
j) One useful strategy to test the quality of graphics presented on the computer is to walk six feet away and see whether the screen is clear from that distance.
k) Lecturers are suggested to dress in off-white, gray, or blue color. Colors such as white, red, or dark green is discouraged (they are too bright or too dark for video shooting). Clothes with stripe patterns are fine.

For further information on the rating system, please see the UIUC EPT holistic rating scales. There are two scales, each called "benchmarks": one for graduate students and one for undergrads.

Appendix 3: UIUC ESL Placement Test Video-Reading Essay Exam: Rating Benchmarks as of February 1998.

Benchmarks for EPT composition scoring
(apply to graduate students)

updated 12/12/96

Too low (effectively places into ESL 400 [formerly 109])

--> insufficient length
--> extremely bad grammar[1]
--> doesn't write on assigned topic; doesn't use any information from the sources[2]
--> majority of essay directly copied[3]
--> summary of source content marked by inaccuracies[4]

400 [formerly 109]

--> dropped sentences and paragraphs
--> whole essay doesn't make sense; hard to follow the ideas
--> poor choice of words
--> lack of cohesion at the paragraph level
--> grammatical/lexical errors impede understanding[1]
--> only summary/restatement of information in same order as source[2]
--> only uses article; no reference to information in video[2]
--> overt plagiarism: direct copying of passages[3]
--> poor understanding of source content
--> summary of source content contains inaccuracies, both major (concepts) and minor (details)[4]
--> insufficient length

401 [formerly 111]

--> reasonable attempt at introduction, body, conclusion
--> has a main idea; more than restatement of article/video[2]
--> lack of cohesion at the essay level
--> some grammatical/lexical errors; essay still comprehensible[1]
--> summarizes/integrates information from both sources[2]
--> covert plagiarism: some attempts at paraphrasing[3]
--> summary of source content may contain a few minor (details) inaccuracies[4]

Exempt, or recommends 402/403 sequence [formerly 400/401]

--> excellent introduction, body, conclusion
--> cohesion at essay level
--> writing flows smoothly
--> grammatical/lexical errors do not impede understanding[1]
--> uses information from both sources to effectively argue thesis (explicit or implied)[2]
--> no or minimal plagiarism (citation of source is desirable, but not necessary)[3]
--> summarization of source content should contain no major (concepts) inaccuracies[4]

Explication and Rationale for Benchmarks

In general, the task of the rater is to evaluate the student's ability to write an essay and use/synthesize source material.

NOTE: No single issue should be used to place a student in a certain level. A combination of factors should be used.

(1) Grammatical & Lexical Errors

Too Low:	Extremely bad grammar; totally incomprehensible.
400:	Grammatical/lexical errors impede understanding. Even after rereading, confused about what student means.
401:	Some grammatical/lexical errors, but essay is still comprehensible. Might have to reread some sentences, but after rereading can basically understand what student means.
Exempt:	Grammatical/lexical errors do not impede understanding. Don't have to stop and reread parts to understand the essay. Types of errors few and tend to be those easily corrected.

(2) Responding to Prompt

Too Low:	Student writes about a topic other than assigned topic or fails to use any source of information. This demonstrates student either doesn't understand instructions or content or is unable to respond to an assigned topic. If student fails to use sources, raters cannot evaluate student's ability to use sources.
400:	Student uses some information from sources, but writes on a topic unrelated or only remotely related to assigned topic (see above rationale). Student simply summarizes/repeats information in the order in which it originally appeared in the source(s), especially if student uses only the article as a source of information. In the case of repeating information in order, rater cannot determine student's ability to organize writing at either paragraph or essay level. In the case of student only using the article, it is difficult to evaluate the student's listening comprehension and ability to use and synthesize information from an aural source. Since 400 gives special attention to listening skills, students who fail to demonstrate listening ability should be placed in 400.
401:	Student summarizes and/or integrates information from both aural and written sources. Student develops a main idea related to the topic and supports it with the sources. There may still be lack of cohesion at the essay level.
Exempt:	Student skillfully uses information from both sources to effectively argue his/her thesis (explicit or implied).

(3) Plagiarism

Too Low:	Majority of essay is directly copied from sources without citation.
400:	Essay contains overt plagiarism. Direct copying of passages/sentences doesn't allow rater to know student's true writing ability. When plagiarism is significant enough to hinder rater's ability to judge how well the student

writes, he/she should be placed in 400. (Guideline: Therefore, if someone directly copies approximately 1/3 of the essay, he/she would be in 400.)

401: Essay contains covert plagiarism. There may be imperfect attempts at summarizing and paraphrasing and isolated incidents of direct copying of no more than a couple sentences. If the rest of the essay demonstrates student's writing is good, student can learn about plagiarism in 401.

Exempt: No or minimal plagiarism. Citation of source is desirable but not necessary.

(4) Accuracy of Content

Too Low: Summary of source content is marked by inaccuracies.

400: Summary of source content contains inaccuracies, both major (concepts) and minor (details).

401: Summary of source content should, on the whole, be accurate but may contain a few minor (details) inaccuracies.

Exempt: Summarization of source content should contain no major (concepts) inaccuracies.

Too low (student should be placed into 113)

--> length insufficient to evaluate
--> no organization of ideas
--> content marked by inaccuracies of source information
--> grammatical and lexical errors impede understanding
--> sentence variety and complexity not present

113L

--> length insufficient to express main idea
--> whole essay does not make sense; difficult to follow ideas
--> paragraph structure not mastered; lack of main idea, focus, cohesion
--> essay lacks a central idea
--> summarizes/restates of sources rather than uses them to support ideas
--> apparent misunderstanding of source material
--> some overt plagiarism
--> some grammatical/lexical errors impede understanding
--> little sentence variety; sentence complexity not mastered

113U

--> whole essay does not make sense; difficult to follow ideas
--> paragraph structure not mastered; lack of main idea, focus, cohesion
--> essay lacks a central idea
--> summarizes/restates of sources rather than uses them to support ideas
--> apparent misunderstanding of source material
--> some overt plagiarism
--> some grammatical/lexical errors impede understanding
--> little sentence variety; attempts at sentence complexity lead to misunderstanding

***_Should source use be included in 113? We have only addressed source understanding_

114

--> LENGTH SUFFICIENT ENOUGH FOR FULL EXPRESSION OF IDEAS
--> ELEMENTS OF ESSAY ORGANIZATION ATTEMPTED; INTRO BODY CONCLUSION
--> paragraph structure mastered
--> attempted use of transitions; some inaccuracies
--> attempt to advance a main idea
--> adequate use of oral and written sources to advance main idea
--> use of oral and written sources demonstrates an understanding of source material
--> covert plagiarism; attempted summary and paraphrase and ISOLATED instances of direct copying
--> some grammatical/lexical errors; essay still comprehensible
--> good sentence variety; some complexity mastered

Appendix 4: UIUC Componential Composition Evaluation Guide and Essay Levels (used for progress measurement during the ESL service courses at UIUC)

[20 Feb 1998]

UIUC ESL Service Courses
Composition Evaluation Guide and Essay Level

	(1)	(2)	(3)	(4)	(5)
FEATURES	FAIL	NARROW FAIL	NARROW PASS	PASS	HIGH PASS
ORGANIZATION					
Degree to which logical flow of ideas and explicitness of plan (intro, thesis, conclusion) are clear and connected	No plan; insufficient length to ascertain organization	Attempted plan is noticeable; inadequate paragraphing	Plan is clear; some cohesion and coherence	Plan is clear; most points connected; coherent; various use of cohesive devices	Sophisticated use of format elements with all points connected and signified with transitions and/or other cohesive devices.
CONTENT					
Degree to which main points are elaborated and/or explained by evidence and detailed reasons	No support; insufficient length and/or extensive plagiarized passages; too short to evaluate	Attempted elaboration; may be a list of related specifics; some plagiarism; inadequate length	Some points elaborated; some acknowledgement of sources	Most points elaborated; argument is developed in a clear and logical manner; characterized by adequate generalizations	All major points elaborated; argument is developed in a clear and logical manner with proper use of sources and leads to sophisticated generalizations and conclusions

BEST COPY AVAILABLE

	(1)	(2)	(3)	(4)	(5)
FEATURES (cont'd)	FAIL	NARROW FAIL	NARROW PASS	PASS	HIGH PASS
CONVENTIONS					
Use of grammatical conventions of standard English (usage, sentence construction, spelling, capitalization, paragraph format)	Many errors; cannot read; confused meaning, problems with sentence construction; insufficient length to evaluate	Some major and many minor errors; confusion; sentence construction not mastered	Developed; few major errors, some minor; meaning unimpaired; mastery of sentence construction	A few minor errors, but no more than one major error	No major errors; one or two minor errors
VOCABULARY AND STYLE					
Degree to which student has mastered register, sentence variation, and word/idiom use including word form	Very poor; essentially translation; sentence variety not mastered	Limited range; frequent errors of word/idiom choice, form & usage that confuse meaning; some sentence variety	Adequate range of word/idiom choice and sentence variety; some errors that do not obscure meaning	Few word form errors, occasionally misused word/idioms; appropriate register; varied sentence structure	Sophisticated range, effective word/idiom choice & usage; word form mastery; register appropriate; varied sentence structure

BEST COPY AVAILABLE

Appendix 5: UIUC ESL Placement Test: Plagiarism Test Specification

EPT Multiple-Choice Test of Awareness of Plagiarism

Origin/History: Early genesis of this spec traces to some test development groups in EIL 360, the UIUC MATESL language testing course. Additional work was done by various independent parties, including: Stacia Steward, Mi-Ok Kim, and Yeonsuk Cho. As of this writing (February 1998), we have just run a trial caboose of plagiarism items with the most recent EPT intake group. We wish to emphasize that this plagiarism spec is still in a draft *form. Despite the extensive work done to date, much needs to be resolved regarding its* internal structure. The VRESSAY given in the next appendix, below, is an example of a much better evolved test specification.

Setting/Mandate: This spec is intended for use in generating multiple-choice items for the English Placement Test for both the undergraduate and graduate sequence ESL service courses at the University of Illinois at Urbana-Champaign. Plagiarism was identified by instructors in the service courses as a useful candidate for phasing into the existing m/c test of English grammar and usage. It should also be noted that plagiarism plays a major deciding role in both the holistic EPT essay rating scale ("benchmarks") and the composition evaluation guide used for progress testing in the service courses.

General Description: The purpose of this spec is to test students' ability to recognize plagiarism using a multiple-choice format that includes several skills involved in writing from sources, such as paraphrasing, quoting, and citing.

Related Specifications, if any: The EPT Specification for Video-Reading Based Academic Essays

Prompt Attributes: The questions pertaining to plagiarism will be preceded by a "mini-lesson" on plagiarism, designed to clarify terminology used in the questions (see sample "mini-lesson" below). The prompt consists of an authentic text (about 100-200 words long). The text should be on a topic of general interest so that students of all academic backgrounds can understand it, but should contain ample "citeworthy" material. Students will read this text, and each item will use a part of the text as it might be used in incorporating the material into academic writing (that is, paraphrased or quoted material with the source cited). Each item may have no more than one mistake in how this material is used, for example a problem with paraphrasing or quoting or citing.

Response Attributes: Students will respond to multiple-choice items which ask whether the source has been used correctly or, if it has been used incorrectly, require them to choose what kind of mistake has been made. Other items will test students' ability to distinguish plagiarized passages from non-plagiarized passages. Distracters should include mistakes that students are likely to make. In order to trial items derived from this spec, it was necessary to compare student performance on those items to performance on the VRESSAY. Hence, a revised componential rating grid (derived from the long-standing progress grid used in ESL courses at UIUC) was developed, one which features an evaluation of plagiarism. This revised grid is given, for the record, in the Specification Supplement below.

Sample "Mini-Lesson"

Plagiarism is considered a very serious matter in the United States. The following questions test your ability to recognize plagiarism and how to use sources correctly in academic writing. For the purpose of these questions, plagiarism is defined as "the unacknowledged use of someone else's idea and/or words (including key words or phrases, as well as longer units like sentences and paragraphs)" in your own writing.

Plagiarism occurs (1) when you borrow an idea or information from another source without acknowledging (giving a citation for) the source; (2) when you do not use quotation marks to

BEST COPY AVAILABLE

show what wording comes exactly from the original source; and (3) when you inadequately paraphrase the original author's ideas. Using much of the author's wording or using the author's sentence structure is considered inadequate paraphrase.

For the purpose of this test, a complete citation should include author's last name, year, and page number. An inaccurate paraphrase or citation is one that contains incorrect information. An inaccurate quotation is one that does not quote the original exactly.

Types of Questions

1. Prompt: paraphrase from the text including parenthetical citation
   Question: The source is used...
   A) correctly.
   B) incorrectly; the paraphrase contains plagiarism.
   C) incorrectly; the paraphrase is inaccurate.
   D) incorrectly; the paraphrase is inaccurate and contains p plagiarism.

2. Prompt: paraphrase or quotation from the text which may or may not include a complete parenthetical citation
   Question: The source is used....
   A) correctly.
   B) incorrectly; the source is not acknowledged.
   C) incorrectly; the reference does not contain enough information.
   D) incorrectly; the information in the reference is inaccurate.

3. Prompt: 1-2 sentence passage that uses quotation marks
   Question: The source is used....
   A) correctly.
   B) incorrectly; the source is not acknowledged.
   C) incorrectly; the quotation marks are misplaced.
   D) incorrectly; the quotation is inaccurate.

4. Prompt: a 1-2 sentence passage that contains plagiarism
   Question: The above passage contains plagiarism. Why?
   A) inadequate paraphrase
   B) no acknowledgment of source
   C) wrong use or placement of quotation marks
   D) All of the above

5. Question: Which of the following paraphrases the passage **best** while still avoiding plagiarism?
   NOTE: (both A&B should include parenthetical citation, order may be reversed)
   A) (an adequately paraphrased passage)
   B) (an inadequately paraphrased passage)
   C) None of the above contain plagiarism.
   D) All of the above contain plagiarism.

(alternatively, A & B could both be inadequate paraphrases OR both be adequate paraphrases)

6. Question: Which of the following **does not** contain plagiarism?

NOTE: (A, B, & C should include parenthetical citation; order may be changed)

A) (an adequately paraphrased passage)
B) (a paraphrase of the same passage which has the same sentence structure as the original, but replaces some words with synonyms)
C) (a paraphrase of the same passage which has a different sentence structure than the original, but which keeps much of the same wording)
D) All of the above contain plagiarism.

alternatively: D) None of the above contain plagiarism.

7. Question: Which of the following uses the source correctly?

NOTE: (A, B, & C should include parenthetical citation; order may be changed)

A) (a passage with quotation marks used correctly)
B) (the same passage with quotation marks misplaced)
C) (the same passage with an inaccurate quotation)
D) All of the above use the source correctly.

alternatively: D) None of the above use the source correctly.

8. Question: Which of the following **does not** contain plagiarism?

NOTE: (order of A, B, & C may be changed)

A) (a paraphrased or quoted passage with proper citation)
B) (the same paraphrased or quoted passage with no citation)
C) (the same paraphrased or quoted passage with an incomplete citation)
D) All of the above contain plagiarism.

alternatively: D) None of the above contain plagiarism.

Sample Items:

(All the items which follow are based on this text. Clearly not all items could be used on the same test since there is considerable overlap, and some items give away the answer to other items.)

"Population Growth"

According to the United Nations Fund for Population Activities, a research and study arm of the UN, the world population growth rate dropped from 1.99 percent in 1960-65 to 1.72 percent in 1975-80. That may not seem significant, but it represents the difference between a world population of 10.5 billion by the year 2110 and a population of 14 billion by that same date--a 20 percent difference.

The numbers may vary, but the implication of either set of figures is the same--population control remains a matter of critical importance to the world. Impressive gains have been made, to be sure: just 10 years ago, for example, only 60 Third World governments supported family planning. Today, 94 of the 124 Third World nations encouraged such planning, the majority of them funding and operating their own programs. In industrialized nations, the changing status of women is credited with having produced a decline in the overall birthrate.

Even so, it is not enough. Population growth on the scale now being predicted promises major food, water and energy shortages, more crime, less freedom and a lower standard of living than is today the case. Indeed, there is barely time enough to fully define and comprehend the problem, much less arrive at consensus solutions by the year 2000.

[Excerpted from The Eagle, Bryan College, July 9, 1982, p.8.]

Type 1 Sample Items:

1. The United Nations Fund for Population Activities reported that the population growth rate of the world was 1.99 percent in 1960-65 but declined in 1975-80 to 1.72 percent (The Eagle, 1982, p.8).

The source is used...

*A) correctly.
B) incorrectly; the paraphrase contains plagiarism.
C) incorrectly; the paraphrase is inaccurate.
D) incorrectly; the paraphrase is inaccurate and contains plagiarism.

2. According to The Eagle (1982), just 10 years ago in 1972, only 60 Third World governments supported family planning, but in 1982, 94 of 124 Third World nations encouraged planning, mostly funding and operating their own programs (p.8).

The source is used...

A) correctly.
*B) incorrectly; the paraphrase contains plagiarism.
C) incorrectly; the paraphrase is inaccurate.
D) incorrectly; the paraphrase is inaccurate and contains plagiarism.

3. In 1972 the governments of 94 Third World countries promoted family planning, while in 1982 60 of them promoted it (The Eagle, 1982, p.8).

The source is used...

A) correctly.
B) incorrectly; the paraphrase contains plagiarism.
*C) incorrectly; the paraphrase is inaccurate.
D) incorrectly; the paraphrase is inaccurate and contains plagiarism.

Type 2 Sample Items:

4. The United Nations Fund for Population Activities reported that the population growth rate of the world was 1.99 percent in 1960-65 but declined in 1975-80 to 1.72 percent (The Eagle, 1982, p.8).

The source is used....

*A) correctly.
B) incorrectly; the source is not acknowledged.
C) incorrectly; the reference does not contain enough information.
D) incorrectly; the information in the reference is inaccurate.

5. The United Nations Fund for Population Activities reported that the population growth rate of the world was 1.99 percent in 1960-65 but declined in 1975-80 to 1.72 percent.

The source is used....

A) correctly.
*B) incorrectly; the source is not acknowledged.
C) incorrectly; the reference does not contain enough information.
D) incorrectly; the information in the reference is inaccurate.

6. The United Nations Fund for Population Activities reported that the population growth rate of the world was 1.99 percent in 1960-65 but declined in 1975-80 to 1.72 percent (1982 newspaper).

The source is used....

A) correctly.
B) incorrectly; the source is not acknowledged.
*C) incorrectly; the reference does not contain enough information.
D) incorrectly; the information in the reference is inaccurate.

7. The United Nations Fund for Population Activities reported that the population growth rate of the world was 1.99 percent in 1960-65 but declined in 1975-80 to 1.72 percent (The Eagle, 1992, p.8).

The source is used....

A) correctly.
B) incorrectly; the source is not acknowledged.
C) incorrectly; the reference does not contain enough information.
*D) incorrectly; the information in the reference is inaccurate.

Type 3 Sample Items

8. According to The Eagle (1982), "population growth on the scale now being predicted promises major food, water and energy shortages, more crime, less freedom and a lower standard of living than is today the case" (p.8).

The source is used....

*A) correctly.
B) incorrectly; the source is not acknowledged.
C) incorrectly; the quotation marks are misplaced.
D) incorrectly; the quotation is inaccurate.

9. Some are claiming "population growth on the scale now being predicted promises major food, water and energy shortages, more crime, less freedom and a lower standard of living than is today the case" (The Eagle, 1982, p.8).

The source is used....

*A) correctly.
B) incorrectly; the source is not acknowledged.
C) incorrectly; the quotation marks are misplaced.
D) incorrectly; the quotation is inaccurate.

10. Some are claiming, "population growth on the scale now being predicted promises major food, water and energy shortages, more crime, less freedom and a lower standard of living than is today the case."

The source is used....

A) correctly.
*B) incorrectly; the source is not acknowledged.
C) incorrectly; the quotation marks are misplaced.
D) incorrectly; the quotation is inaccurate.

11. According to The Eagle (1982), population growth on the scale now being predicted promises "major food, water and energy shortages, more crime, less freedom and a lower standard of living than is today the case" (p.8).

The source is used....

A) correctly.
B) incorrectly; the source is not acknowledged.
*C) incorrectly; the quotation marks are misplaced.
D) incorrectly; the quotation is inaccurate.

12. According to The Eagle (1982), "population growth promises major food, energy and water shortages, more crime, less freedom and a lower standard of living than is today the case" (p.8).

The source is used....

A) correctly.
B) incorrectly; the source is not acknowledged.
C) incorrectly; the quotation marks are misplaced.
*D) incorrectly; the quotation is inaccurate.

Type 4 Sample Items

13. According to The Eagle (1982), just 10 years ago in 1972, only 60 Third World governments supported family planning, but in 1982, 94 of 124 Third World nations encouraged planning, mostly funding and operating their own programs (p.8).

The above passage contains plagiarism. Why?

*A) inadequate paraphrase
B) no acknowledgment of source
C) wrong use or placement of quotation marks
D) All of the above

14. Some have claimed that the population growth rate of the world was 1.99 percent in 1960-65 but declined in 1975-80 to 1.72 percent.

The above passage contains plagiarism. Why?

A) inadequate paraphrase
*B) no acknowledgment of source
C) wrong use or placement of quotation marks
D) All of the above

15. Population growth continues at a rate which will cause serious problems. Time is short. In fact, the year 2000 is approaching so soon that there is hardly time to define, understand, and reach solutions.

The above passage contains plagiarism. Why?

A) inadequate paraphrase
*B) no acknowledgment of source
C) wrong use or placement of quotation marks
D) All of the above

16. According to The Eagle (1982), population growth on the scale now being predicted promises "major food, water and energy shortages, more crime, less freedom and a lower standard of living than is today the case" (p.8).

The above passage contains plagiarism. Why?

A) inadequate paraphrase
B) no acknowledgment of source
*C) wrong use or placement of quotation marks
D) All of the above

Type 5 Sample Item

17. Which of the following paraphrases the passage **best** while still avoiding plagiarism?

*A) The Eagle (1982) has argued that the issue of controlling the world's population has serious implications for the world (p.8).
B) The Eagle (1982) has argued that population control is a matter of critical importance in the world (p.8).
C) None of the above contain plagiarism.
D) All of the above contain plagiarism.

Type 6 Sample Item

18. Which of the following **does not** contain plagiarism?

*A) According to The Eagle (1982), population growth continues at a rate which will cause serious problems. Time is short. In fact, the year 2000 is approaching so soon that there is hardly time to define, understand, and reach solutions (p.8).

B) According to The Eagle (1982), there is hardly time enough to completely describe and understand the problem of population growth, much less reach agreement on solutions by the year 2000 (p.8).

C) According to The Eagle (1982), the year 2000 is approaching so soon that there is barely enough time to arrive at consensus solutions or to fully comprehend and define the problem of population growth (p.8).

D) All of the above contain plagiarism.

Type 7 Sample Items

19. Which of the following uses the source correctly?

*A) According to The Eagle (1982), "population growth on the scale now being predicted promises major food, water and energy shortages, more crime, less freedom and a lower standard of living than is today the case" (p.8).

B) According to The Eagle (1982), population growth promises "major food, water and energy shortages, more crime, less freedom and a lower standard of living" than is today the case (p.8).

C) According to The Eagle (1982), "population growth promises major food, energy and water shortages, more crime, less freedom and a lower standard of living than is today the case" (p.8).

D) None of the above contain plagiarism.

20. Which of the following uses the source correctly? OR ALTERNATIVELY Which of the following **does not** contain plagiarism?

A) The United Nations Fund for Population Activities reported that "the world population growth rate dropped from 1.99 percent in 1960-65 to 1.72 percent in 1975-80" (The Eagle, 1982, p.8).

B) The United Nations Fund for Population Activities reported that the growth rate of the world's population "dropped from 1.99 percent in 1960-65 to 1.72 percent in 1975-80" (The Eagle, 1982, p.8).

C) The United Nations Fund for Population Activities reported that the rate of population growth in the world "dropped from 1.99 percent in 1960-65 to 1.72 percent in 1975-80" (The Eagle, 1982, p.8).

*D) None of the above contain plagiarism.

Type 8 Sample Item

21. Which of the following **does not** contain plagiarism?

*A) The United Nations Fund for Population Activities reported that the population growth rate of the world was 1.99 percent in 1960-65 but declined in 1975-80 to 1.72 percent (The Eagle, 1982, p.8).

B) The United Nations Fund for Population Activities reported that the population growth rate of the world was 1.99 percent in 1960-65 but declined in 1975-80 to 1.72 percent.

C) The United Nations Fund for Population Activities reported that the population growth rate of the world was 1.99 percent in 1960-65 but declined in 1975-80 to 1.72 percent (1982 newspaper).

D) None of the above contain plagiarism.

"Jet Lag"

The problem of Jet Lag is one every international traveller comes across at some time. However, the effects of rapid travel on the body are actually far more disturbing than we realize. Jet Lag is not a psychological consequence of having to readjust to a different time zone. It is due to alterations in the body's physiological regulatory mechanisms, specifically the hormonal systems, in a different environment.

Different bodily events are governed by different factors. The hormone cortisol, which controls salt and water excretion, is made in the morning, wherever the body is. But the growth hormone is released during sleep, whenever in the day that sleep occurs. Normally these two hormones are separated by seven or eight hours, but if the body arrives at a destination in the early morning (local) and goes to sleep as soon as possible, the two hormones will be secreted simultaneously.

[Excerpted from "Jet Lag," by Bruce Durie, Saudi Arabian Airlines, June 1985, p. 32.]

1. According to Durie, jet lag is not a psychological consequence of having to readjust to a different time zone; rather it is due to alterations in the body's hormonal systems (Durie, 1985, p.32).

The source is used...

A) correctly.
B) incorrectly; the paraphrase is inaccurate.
*C) incorrectly; the paraphrase contains plagiarism.
D) incorrectly; the paraphrase is inaccurate and contains plagiarism.

2. Jet lag results when the growth hormone, normally secreted in the morning, and the cortisol hormone, secreted at night, are released simultaneously (Durie, 1985, p.32).

The source is used...

A) correctly.
B) incorrectly; the paraphrase contains plagiarism.
*C) incorrectly; the paraphrase is inaccurate.
D) incorrectly; the paraphrase is inaccurate and contains plagiarism.

3. According to Durie, most people who have travelled internationally have also experienced Jet Lag (Durie, 1985, p.32).

The source is used...

A) incorrectly; the paraphrase is inaccurate.
B) incorrectly; the paraphrase contains plagiarism.
*C) correctly.
D) incorrectly; the paraphrase is inaccurate and contains plagiarism.

4. According to Durie, "jet lag" is not a psychological consequence of having to readjust to a different time zone; rather it is due to alterations in the body's hormonal systems (Durie, 1985, p.32).

The above passage contains plagiarism. Why?

A) wrong use or placement of quotation marks

B) no acknowledgement of source
*C) inadequate paraphrase
D) All of the above

5. Someone argues that jet lag results when the cortisol hormone, normally secreted in the morning, and the growth hormone, secreted at night, are released simultaneously.

The above passage contains plagiarism. Why?

A) wrong use or placement of quotation marks
B) inadequate paraphrase
*C) no acknowledgement of source
D) All of the above

6. According to Durie, jet lag is not a "psychological consequence of having to readjust to a different time zone; rather it is due to alterations in the body's hormonal systems" (Durie, 1985, p.32).

The above passage contains plagiarism. Why?

A) inadequate paraphrase
B) no acknowledgement of source
*C) wrong use or placement of quotation marks
D) All of the above

Specification Supplement

Additional Passages Which Might be Used in Generating Items

"Euro Disney: Will it Survive?"

The massive Euro Disney resort rising out of old sugar-beet fields 30 km east of Paris celebrated its first anniversary in mid-April with a burst of fireworks. But this fairy tale is still far from a happy ending. For investors, Euro Disney has turned out to be the financial equivalent of the roller coaster that roars through Big Thunder Mountain--a terrifying plunge from visions of windfall profits to the present reality of deepening losses.

The problem for Euro Disney executives is that too many visitors are daytrippers from Paris or short-term guests. Surrounding the theme park is a massive $5.2 billion resort complex that includes 5,700 hotel rooms. But hotel bookings have fallen far short of predictions, and analysts say Euro Disney will lose as much as $230 million in its first year. Its shares, which hit a high of $37 last spring, are now worth just $20. One French bank, Paribas, maintains that Euro Disney faces years of losses, and has urged stockholders to sell now. European families may have embraced the Disney dream. But investors in the fairy tale are still searching for signs that they will live happily--and lucratively--ever after.

[Excerpted from "Where's the Magic?" by Andrew Phillips, Maclean's, May 3, 1993, p. 47.]

"Americanization in Mexico"

Although U.S. influence has been felt in Mexico for decades, most agree that it has become more visible since President Carlos Salinas de Gortari took office in 1988. Mr. Salinas opened the doors to U.S. goods in an effort to drive down domestic prices and to make Mexican industries more competitive.

The flow of U.S. products soon turned into an avalanche. Mexico City's supermarkets were flooded with U.S. cereals, canned vegetables, detergents and almost any product available in the United States. Mr. Salinas later authorized U.S. companies to open franchises in Mexico. Large neon signs of McDonald's, Burger King, Arby's Kentucky Fried

BEST COPY AVAILABLE

Chicken, Subway and Domino's Pizza began to pop up along the main streets of Mexican cities.

Carlos Monsivais, one of Mexico's most respected left-of-center social analysts, said U.S. influence is going far beyond consumption habits to affecting Mexican traditions and even the Spanish language.

[Excerpted from "Some Mexicans Fear Growing 'Gringoization' of Their Country," by Andres Oppenheimer, The Journal of Commerce, October 19, 1993, p. 10A.]

"Sustainable Agriculture"

For nearly four decades after World War II, U.S. agriculture was the envy of the world, almost annually setting new records in crop production and labor efficiency. During this period U.S. farms became highly mechanized and specialized, as well as heavily dependent on fossil fuels, borrowed capital and chemical fertilizers and pesticides. Today the same farms are associated with declining soil productivity, deteriorating environmental quality, reduced profitability and threats to human and animal health.

A growing cross section of American society is questioning the environmental, economic and social impacts of conventional agriculture. Consequently, many individuals are seeking alternative practices that would make agriculture more sustainable.

Sustainable agriculture addresses many serious problems afflicting U.S. and world food production: high energy costs, ground water contamination, soil erosion, loss of productivity, depletion of fossil resources, low farm incomes and risks to human health and wildlife habitats. It is not so much a specific farming strategy as it is a system-level approach to understanding the complex interactions within agricultural ecologies.

[Excerpted from "Sustainable Agriculture," by J. Reganold, R. Papendick, and J. Parr, Scientific American, June 1990, p. 112.]

"Westernizing Oriental Eyes"

"Thanks to increased immigration of Asians to the United States, as well as a flood of American videos, movies, and fashion magazines into the Far East, more and more Asians identify with Western features--particularly big, beautiful eyes," says Beverly Hills plastic surgeon Ronald Matsunaga. And thanks to a cosmetic-surgery technique that Matsunaga calls a "Westernization procedure," Oriental eyes can now be molded into an Occidental look.

There are two distinct differences between Asian and Caucasian eyes, Matsunaga points out. The upper eyelid in Asians lacks a fold, and in 85 percent of Oriental people, there is a webbing of skin on the nasal side of the eye.

"In the past, surgeons dealt only with adding a fold to make the eye look bigger and more aesthetically pleasing. But if the web was removed, it frequently grew back. My technique, however, removes the web permanently," says Matsunaga, who teaches facial plastic surgery at the University of Southern California at Los Angeles.

[Excerpted from "Westernizing Oriental Eyes," by Sherry Baker, Omni, October 1985, p. 46.]

Additional Comments on text selection:

All of the sample passages in this spec are excerpted from much longer articles which have been used in the ESL Service Courses at some time. In excerpting, efforts were made to "create" a text which was short (100-200 words) and yet cohesive and one which seemed fairly "complete" by itself. For this reason, it was not always possible to excerpt a few paragraphs of contiguous text. Rather, for some of the passages, excerpts come from different parts of the article (compare original Jet Lag article which is attached to excerpt in this spec).

As specified in the Prompt Attributes, efforts were made to select and excerpt texts which contain "citeworthy material." Different types of text material are "quoteworthy" and "citeworthy" for different purposes. Therefore, an attempt was made in selecting the additional passages listed at the end of the spec to include passages which contain a variety of types of "citeworthy" material:

numbers/statistics -- Population Growth, Euro Disney
analogy -- Euro Disney (roller coaster)
scientific/biological characteristics or processes -- Jet Lag, Sustainable Agriculture, Westernizing Oriental Eyes -- sort of
historical progression -- Sustainable Agriculture
experts being quoted directly -- Westernizing Oriental Eyes
or expert opinions reported -- Americanization in Mexico
quotable author's position or claim --
    Population Growth (2nd paragraph, 1st sentence; 3rd paragraph)
    Jet Lag (1st paragraph, 3rd & 4th sentences)
    Sustainable Agriculture (2nd paragraph; 3rd paragraph, last sentence)
    (many other types of citeworthy material exist; this is simply a sample)

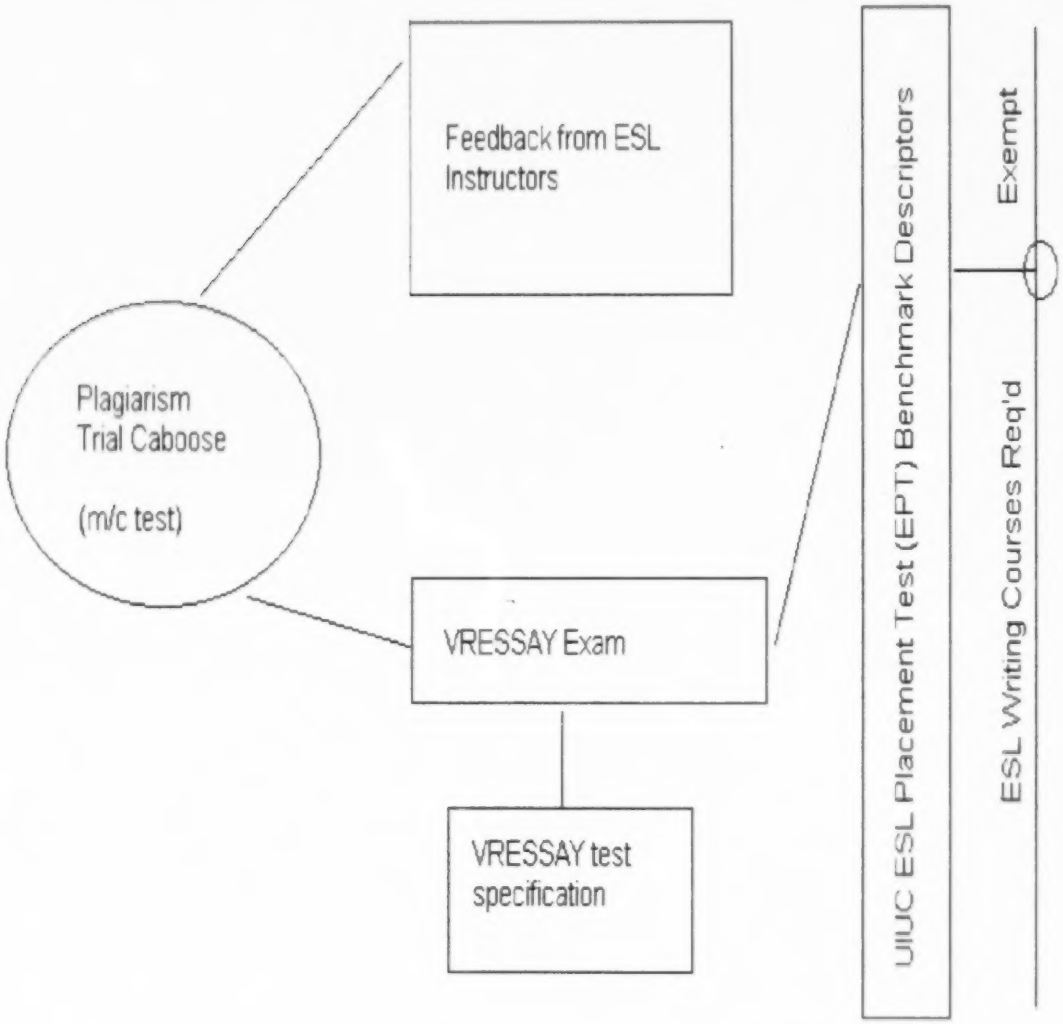
Of course, whether all these passages are appropriate and which ones are best remains open to debate.

VRESSAY/Plagiarism Caboose (January 1998) Componential Rating Rubric

In January 1998, a trial "caboose" was added to the EPT at UIUC. This caboose was formed of items developed from this specification. In order to fully analyze those items, it was necessary to compare their results to the VRESSAY. Operational rating of the VRESSAY is holistic rather than componential and so detailed comparison of item performance to source use (in the essays) was not feasible. Furthermore, the longstanding ESL progress rating grid (used in the service courses) did not detail plagiarism as a component to sufficient detail, and so a new grid was developed by which to rate the VRESSAY and permit criterion data to performance on the caboose items. Following, for the record, is that grid:

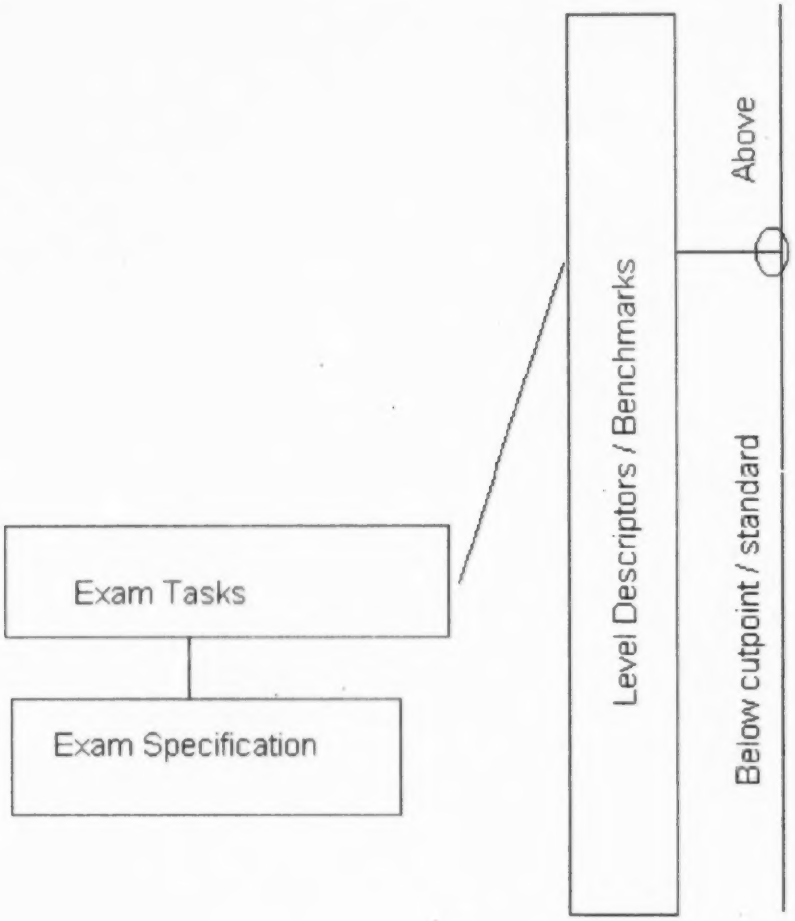
Features	1	2	3	4
Organization	No plan; insufficient to ascertain organization	No clear plan or does not follow plan; lack of paragraph and essay cohesion	Noticeable plan. Reasonable attempt at introduction, body, conclusion; some paragraph and essay cohesion.	Clear plan; excellent introduction, body, conclusion; cohesion at paragraph and essay levels.
Content	No support or elaboration of ideas; insufficient length to evaluate. Irrelevant to assigned topic; fails to use any information from sources; or summary of source content marked by inaccuracies.	Attempted elaboration; minimal or ineffective support for ideas, or insufficient length. Some source use, but may be insufficient to evaluate source understanding or contains major (concepts) and minor (details) inaccuracies.	Most points elaborated. Ideas developed and supported with the sources; overall accurate, but may contain a few minor (details) inaccuracies.	All major points elaborated. Ideas developed, and effective use of (both) sources to support ideas. No major (concepts) inaccuracies in summary of source content.
Plagiarism *Note: this feature is applicable only if sources are used or if there is sufficient evidence of source use to evaluate. Otherwise, put "NA" instead of points.	Majority of essay is directly copied from sources with or without citation.	Overt plagiarism; direct copying hinders rater's ability to evaluate student's true writing ability; 1/3 of the essay contains directly copied phrases / sentences; may cite sources, but not necessary.	Covert plagiarism; inadequate summary and paraphrase; no more than a couple of directly copied sentences; may cite sources, but not necessary.	No or minimal plagiarism; effective paraphrase; citation of source is desirable but not necessary.
Linguistic expression	Extremely bad grammar; totally incomprehensible. No sentence variety or complexity.	Grammatical / lexical errors frequent and impede understanding; awkwardness. Little sentence variety and sentence complexity not mastered.	Some grammatical / lexical errors, but still comprehensible; some sentence variety; some complexity.	Grammatical / lexical errors do not impede understanding. Few errors, mostly easily corrected; sentence variety; sophisticated vocabulary usage and sentence complexity mastered.

Figure 1: the relationship of the experimental plagiarism caboose to the UIUC EPT benchmarks and exemption standard/decision



BEST COPY AVAILABLE

Figure 2: the generalized relationship of a criterion-referenced test and its specification to an external cutpoint / standard



BEST COPY AVAILABLE